Hotel Booking Cancelation Classification

By: Sarah Alabdulwahab & Asma Althakafi

Hotel Bookings Data

The dataset contains **32 features** and **119,390 observations**

o Hotel Type

o Arrival Date

o Stays

o Distribution Channel

o Meal Package

o Country

o Previous Cancellations

o Number of People

Is Canceled – The Target

GOAL

Classify the customers that will **cancel** their hotel booking and those that will not

Cleaning Process

Drop
Duplicates

Drop
Nulls

Drop & merge
Features

Remove
Outliers

Value
Conversion

After cleaning, the dataset contains **25 features** and **81,229 observations**

Which Values Did We Convert?

Country

ISO 3166[1]			ISO 3166-1[2]			ISO 3166-2[3]	
Country name[5]	Official state name[6]	Sovereignty[6][7][8]	Alpha-2 code[5]	Alpha-3 code[5]	Numeric code[5]	Subdivision code links[3]	Internet ccTLD[9]
Afghanistan	The Islamic Republic of Afghanistan	UN member state	AF	AFG	004	ISO 3166-2:AF	.af
Akrotiri and Dhekelia – See United Kingdom, The.							
Åland Islands	Åland	Finland	AX	ALA	248	ISO 3166-2:AX	.ax
Albania	The Republic of Albania	UN member state	AL	ALB	008	ISO 3166-2:AL	.al

Scraped from Wikipedia

Visualizations

Moving to Tableau...

Hotel Booking Cancelation Classification

Percentage of Bookings	Percentage of Cancellations	Percentage of Cancellations per Arrival Data

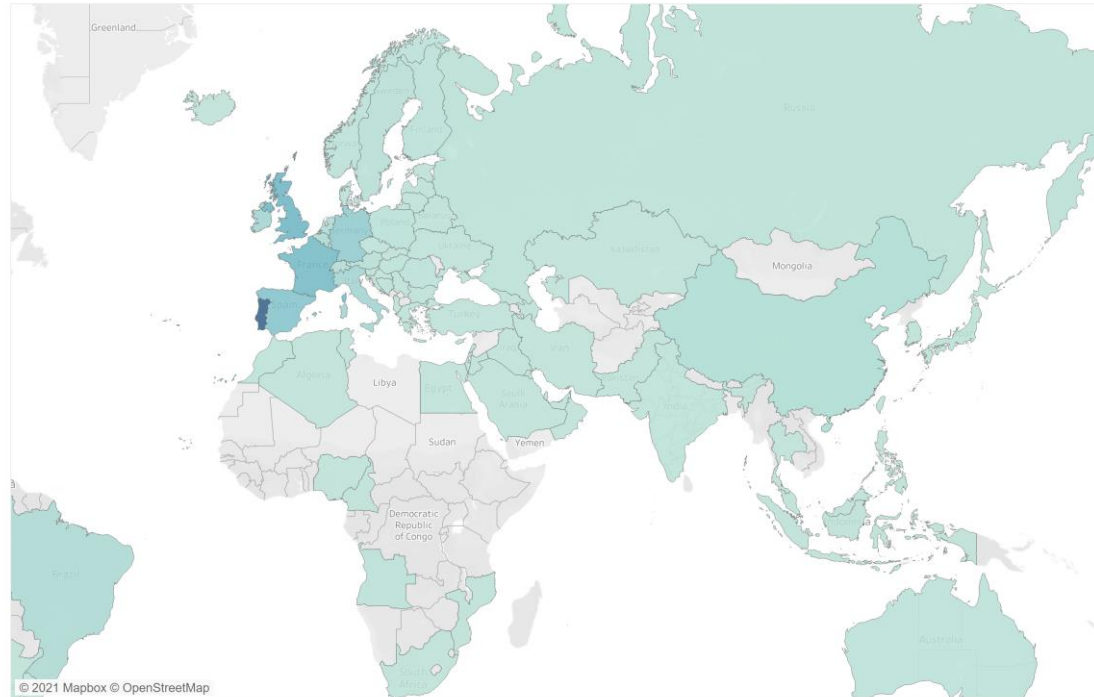## Percentage of Bookings

Country   All

% of Booki..   0.01% ▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓ 31.47%

Percentage of Bookings in each Country

© 2021 Mapbox © OpenStreetMap

Hotel Booking Cancelation Classification

Percentage of Bookings	Percentage of Cancellations	Percentage of Cancellations per Arrival Data

Percentage of Cancellations

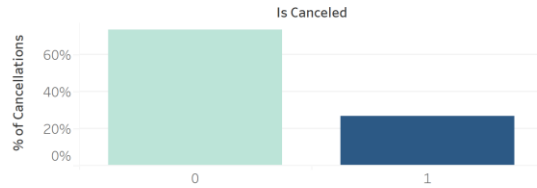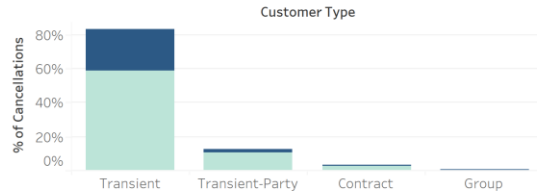| Is Canceled	Hotel	Customer Type	Distribution Channel	Meal	Reserved Room Type
All	All	All	All	All	All

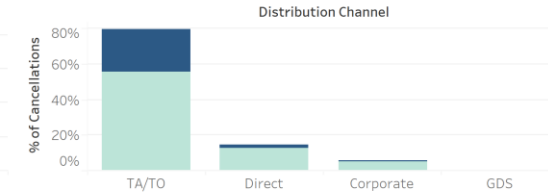Is Canceled ▢ 0   ▉ 1

Percentage of Cancellations
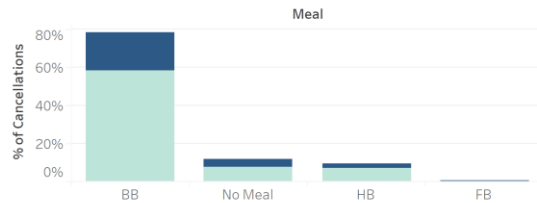
Percentage of Cancellations per Hotel

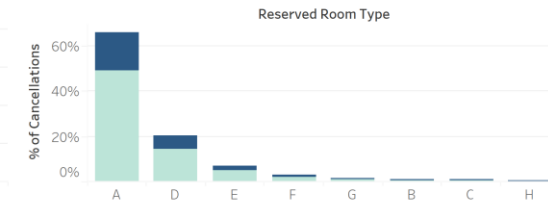### Percentage of Cancellations per Customer Type

Percentage of Cancellations per Distribution Channel

Percentage of Cancellations per Meal

Percentage of Cancellations per Reserved Room Type

Hotel Booking Cancelation Classification

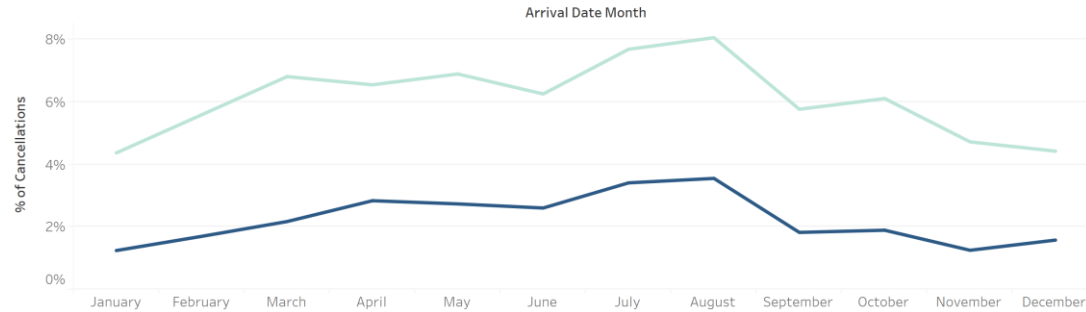| Percentage of Bookings	Percentage of Cancellations	Percentage of Cancellations per Arrival Data

Percentage of Cancellations per Arrival Data

Is Canceled
All

Arrival Date Month
All

Arrival Date Year
All

Is Canceled ■ 1   ■ 0

Percentage of Cancellations per Arrival Data Month

Arrival Date Month

Percentage of Cancellations per Arrival Data Year

Arrival Date Year
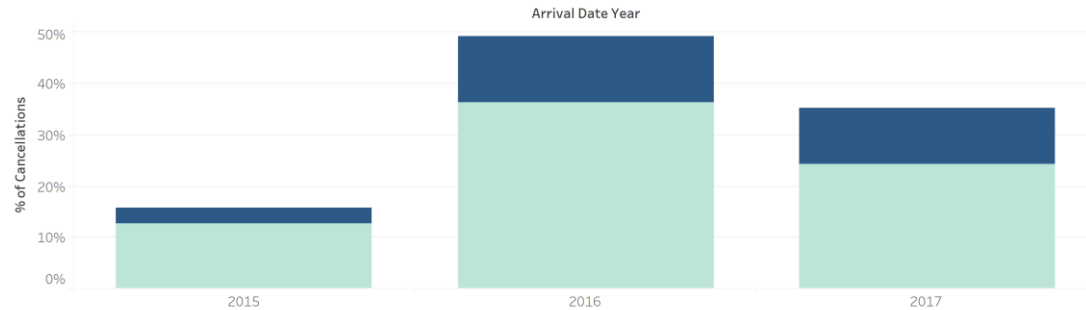
Models

K Neighbors Classifier

Logistic Regression

Decision Tree Classifier

Bernoulli Naive Bayes

Random Forest Classifier

Gradient Boosting Classifier

Bagging with Decision Tree Classifiers

Ada Boost with Decision Tree Classifiers

Experiments

Fit the models on the data with:

1 All observations and features

2 All features and downsampling of the negative observations

3 Selected features based on our interpretation

4 Selected features using Permutation Importance

Experiment 1 – Results

	Accuracy	Precision	Recall	F1
Gradient Boosting	83.96%	83.35%	83.96%	83.37%
Bagging	82.50%	81.77%	82.50%	81.88%
Decision Tree	81.99%	81.39%	81.99%	81.57%

Experiment 2 – Results

	Accuracy	Precision	Recall	F1
Gradient Boosting	81.15%	81.33%	81.15%	81.13%
Bagging	79.83%	80.16%	79.83%	79.78%
Random Forest	79.62%	80.01%	79.62%	79.56%

Experiment 3 – Results

	Accuracy	Precision	Recall	F1
Bagging	75.07%	72.20%	75.07%	71.87%
Gradient Boosting	74.19%	71.21%	74.19%	71.41%
K Neighbors	73.77%	70.88%	73.77%	71.31%

Permutation

Feature importances using Permutation Importance
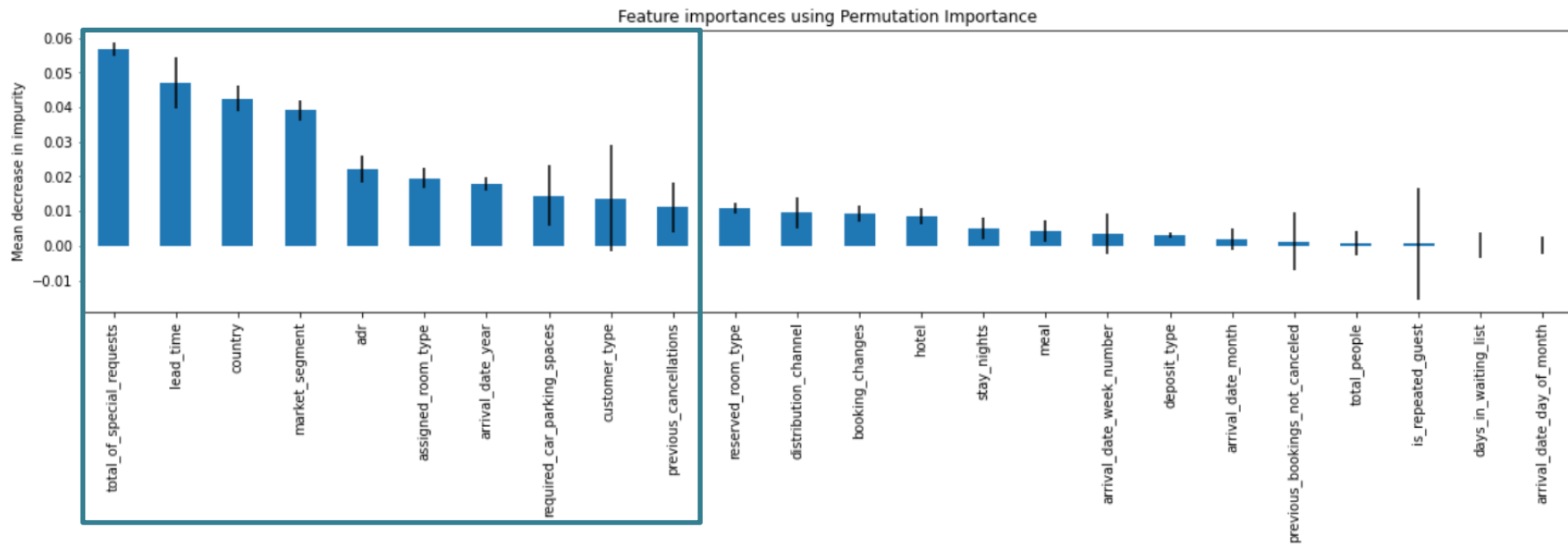
Experiment 4 – Results

	Accuracy	Precision	Recall	F1
Gradient Boosting	81.61%	80.87%	81.61%	81.03%
Bagging	81.10%	80.36%	81.10%	80.55%
Ada Boost	80.33%	79.85%	80.33%	80.04%

Conclusions

- Overall, Gradient Boosting Classifier performed the best, followed by Bagging

- Considering the number of features and the F1 score, the 4th experiment was the best

THANKS!

ANY QUESTIONS?